SINGLE MOLECULE REAL-TIME (SMRT®) SEQUENCING OF FULL-LENGTH HLA-DRB1

Thomas R Turner[1], Cristina Guijarro[1], James Robinson[1,2], Alasdair JM McWhinnie[1], J Alejandro Madrigal[1,2], Neema P Mayor[1,2], Steven GE Marsh[1,2]

1 Anthony Nolan Research Institute, Royal Free Hospital, London UK, **2** UCL Cancer Institute, Royal Free Campus, London UK

INTRODUCTION

- Allelic level HLA typing facilitates optimal donor selection for haematopoetic stem cell transplantation, resulting in improved patient survival and reduced side effects such as GvHD
- Unambiguous, high resolution typing of HLA-DRB1 has been challenging due to a lack of reference sequences (Figure 1), allele length variability and the potential to co-amplify other HLA-DR genes
- Pacific Biosciences' Single Molecule Real Time (SMRT) DNA sequencing allows long-amplicon sequences to be generated in isolation, resolving phase and ambiguities
- We have developed a robust, multiplex method to sequence full-length DRB1 alleles from up to 32 DNA barcoded samples in a single SMRT DNA sequencing reaction
- Single-fragment amplicons contain complete exons 1-6 and introns 1-5, requiring no *in silico* assembly
- Our PCR protocol is DRB1 specific, compatible with template DNA from various sources and amplifies all DRB1 allele groups
- Data presented here demonstrates the ability of SMRT DNA sequencing to provide full-length DRB1 sequences, revealing novel polymorphisms with potential clinical significance or evolutionary interest

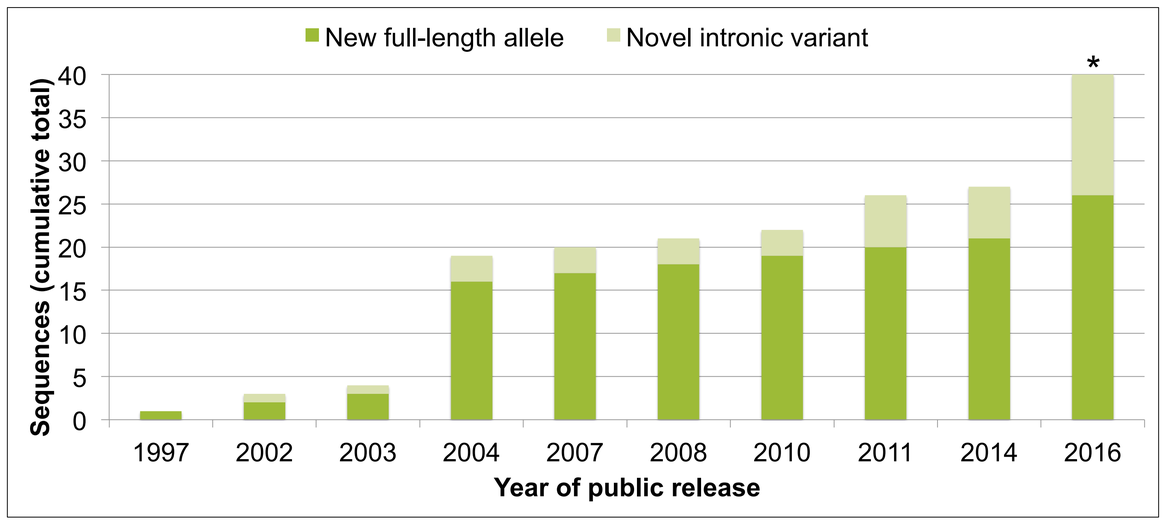

Figure 1: Cumulative total of full-length genomic DRB1 sequences submitted to the IPD-IMGT/HLA Database over time. * The increase in sequences in 2016 is solely due to submissions by Anthony Nolan and is up to date as of April (Release 3.24.0, 2016-04-15).

METHODS

- DNA barcoded primers were designed to bind to the 5´ and 3´ UTRs of DRB1
- Primers were tested by PCR amplification of B-lymphoblastoid cell lines (B-LCLs) homozygous for each DRB1 allele group, including those where full-length genomic sequences were available
- Optimisation of PCR and library preparation was performed to improve yield and allele balance with heterozygous samples
- In-house bioinformatic tools (See Poster 90) have been adapted to deal with variable length sequences
- Novel variants of known alleles were repeat sequenced to confirm polymorphisms
- HLA class I typing was performed (with SMRT DNA sequencing) as required for submission to the IPD-IMGT/HLA Database

RESULTS

- Alleles from each of the 13 DRB1 allele groups were successfully amplified from homozygous B-LCLs, as well as in 23 heterozygous combinations from B-LCLs and other sources
- We generated over 90 full-length genomic sequences from 38 distinct alleles (Figure 2), 13 of which were included in the latest release of the IPD-IMGT/HLA Database (Figure 1)

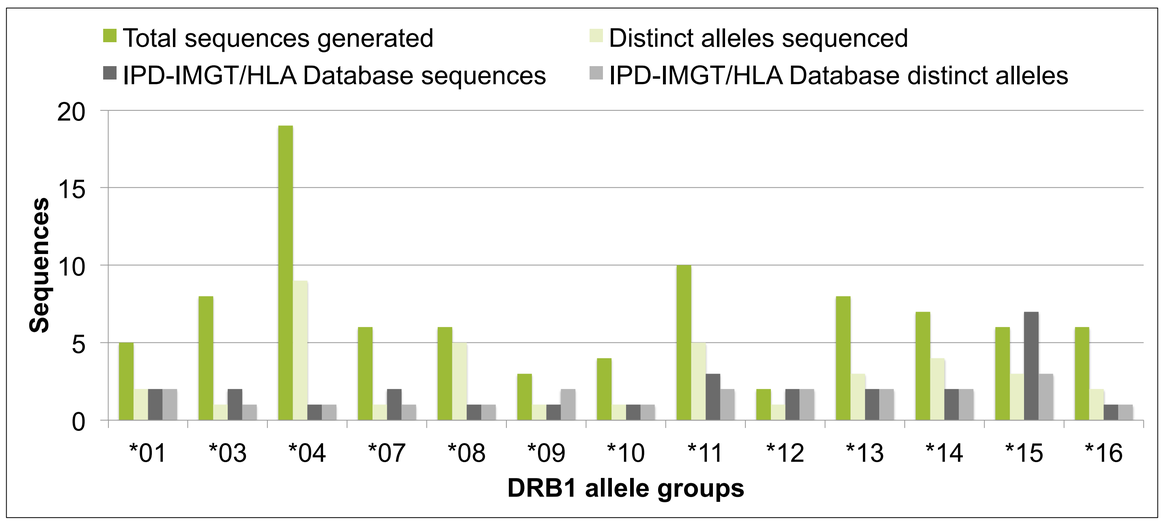

Figure 2: Full-length genomic sequences generated by Anthony Nolan compared to those in the IPD-IMGT/HLA Database prior to our submission of any sequences (Release 3.23.0, 2016-01-19). Several different alleles have been sequenced multiple times from different samples. Many of these require confirmatory sequencing and HLA class I typing before submission to the IPD-IMGT/HLA Database. Several allele groups remain under-represented.

- Widespread intronic variation, particularly in homopolymer and microsatellite regions has been observed, including in samples previously typed as the same allele
- For example, the DRB1*04:01:01 allele from BM14, a B-LCL of southern European origin, contains intronic polymorphisms not present in the same allele from several northern European samples (Figure 3)
- Two full-length genomic sequences in the IPD-IMGT/HLA Database are to be corrected, resolving a homopolymer region in B-LCL PGF and multiple discrepancies in SSTO
- The DRB1*08:01:01 allele from B-LCL MADURA was extended and corrected and found to be identical to an DRB1*08:01:03 allele. The latter allele has now been removed from the IPD-IMGT/HLA Database

```
04:01:01_BM14    CCCTGTGAGTGGAGAATCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT----------GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCCGCCATCTGTGAGCA
04:01:01_BOLETH  CCCTGTGAGCGGAGAATCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT----------------GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCCGCCATCTGTGAGCA
04:01:01_BSM     CCCTGTGAGCGGAGAATCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT----------------GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCCGCCATCTGTGAGCA
04:01:01_MCF     CCCTGTGAGCGGAGAATCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT----------------GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCCGCCATCTGTGAGCA
04:01:01_AN01    CCCTGTGAGCGGAGAATCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT----------------GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCCGCCATCTGTGAGCA
04:01:01_AN02    CCCTGTGAGCGGAGAATCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT----------------GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCCGCCATCTGTGAGCA
04:02:01_YAR     CCCTGTGAGCGGAGAATCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT--GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCCGCCATCTGTGAGCA
04:03:01_SSTO    CCCTGTGAGCGGAGAATCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCCGCCATCTGTGAGCA
04:05:01_LKT3    CCCTGTGAGCGGAGAATCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCCGCCATCTGTGAGCA
04:07:01_JHF     CCCTGTGAGCGGAGAATCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT----------GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCCGCCATCTGTGAGCA
04:11:01_EC      CCCTGTGAGCGGAGAATCTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT----GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCCGCCATCTGTGAGCA
                 ********** ************************************************** ****************************** ***************
```

Figure 3: Alignment of a microsatellite region in intron 2 of DRB1*04 sequences. The number of GT/GA repeats varies between DRB1*04 alleles, but is highly similar between the DRB1*04:01:01 sequences from different samples. The DRB1*04:01:01 sequence from BM14 contains additional polymorphisms (highlighted). Sequences from MCF, AN01 and AN02 will be confirmed by repeat sequencing, which will resolve their microsatellite differences.

CONCLUSIONS

- We have developed a robust, specific and high-throughput method for generating full-length genomic sequences of HLA-DRB1
- Several sequences have been submitted to the IPD-IMGT/HLA Database, and more generated in the past few months than have been submitted in the past 20 years
- We can identify new DRB1 variants, including those with intronic polymorphisms, as well as extend and correct sequences of known alleles
- By sequencing samples of diverse ethnicities, we anticipate the discovery of further novel DRB1 polymorphisms, which may have clinical significance or be of evolutionary interest

Reg charity no 803716/SC038827 978PA/0416

None of the authors have any conflict of interest to declare